Data Loading
and
Data Preprocessing Phase

Phase 3: Development Part 1

Step 1: Loading Libraries and Dataset:

Data set link :

https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset

```
import numpy as np
import pandas as pd
import io
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error,mean_absolute_error
import seaborn as sns
from sklearn.metrics import accuracy_score


from google.colab import files
uploaded=files.upload()


data = pd.read_csv(io.BytesIO(uploaded['MSFT.csv']))
```

```
df=pd.DataFrame(data)
```

Step 2: Exploring the Dataset

```
# Display the first few rows, columns,etc of the dataset to understand its structure
data.head()
data.describe()
data.columns()
print(df.head())
```

Explanation: The dataset was loaded and the first few rows, descriptive statistics, and column names were displayed for initial exploration and to understand its structure.

Step 3: Handling Missing Values:

```
# 3. Dataset Cleaning (Handle Missing Values)
# Check for missing values and handle them if necessary
print("Missing Values:")
print(df.isnull().sum())  # Check for missing values

# If there are missing values, you can choose to fill or drop them as needed.
# For simplicity, let's drop rows with missing values in this example.
df.dropna(inplace=True)
```

Explanation: Identified specific columns with missing values and filled them or dropped them as needed.

Step 4: We save and load the Preprocessed Data

Preprocessed dataset link: MSFT.xlsx

Conclusion:

In this phase, we successfully loaded the dataset, explored its initial structure, handled missing values, and saved the preprocessed data for future analysis. The dataset is now ready for in-depth analysis, modeling, and prediction in the subsequent phases of the project.